PATENT ABSTRACTS OF JAPAN

(11)Publication number : 07-104782
(43)Date of publication of application : 21.04.1995

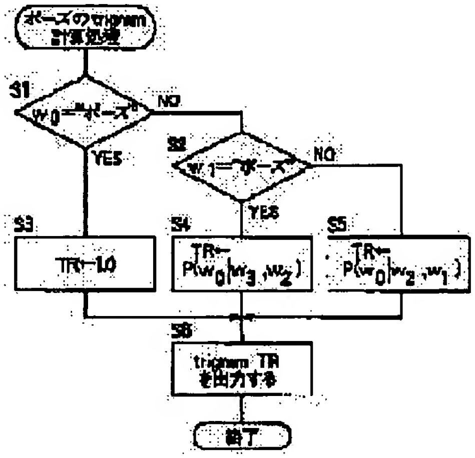
(51)Int.CI. G10L 3/00

(21)Application number : 05-247836
(22)Date of filing : 04.10.1993

(71)Applicant : ATR ONSEI HONYAKU TSUSHIN KENKYUSHO:KK
(72)Inventor : MURAKAMI JINICHI

(54) VOICE RECOGNITION DEVICE

(57)Abstract:
PURPOSE: To provide a voice recognition device which can obtain a high sentence recognition rate as compared with a conventional example even when an inputted spoken voice includes a pause or redundant word.
CONSTITUTION: The voice recognition device which performs voice recognition by calculating a probability value of a word to be recognized in the spoken voice sentence consisting of a character string inputted by referring to a specific statistical language model on the basis of one or plural words connected in front of the word sets the probability value of the language model of the pause in the spoken voice sentence or the word connected to the redundant word to 1 or a value close to 1, and calculates the probability of the language model of the work by skipping the pause or redundant word connected to a work other than the pause or redundant work across the pause or redundant work, thereby performing the voice recognizing process.



LEGAL STATUS

[Date of request for examination]	08.05.2000
[Date of sending the examiner's decision of rejection]	07.01.2003
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]	
[Date of final disposal for application]	
[Patent number]	
[Date of registration]	
[Number of appeal against examiner's decision of rejection]	
[Date of requesting appeal against examiner's decision of rejection]	
[Date of extinction of right]	

* NOTICES *

Japan Patent Office is not responsible for any damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

OPERATION

[Function] While the above-mentioned speech-recognition means makes the probability value of the language model of a word connected to the pause or the redundancy word in the above-mentioned utterance voice sentence the predetermined value which is a value near 1 or 1, when connecting with words other than a pause and a redundancy word through a pause or a redundancy word, it calculates the probability value of the language model of a word by skipping the above-mentioned pause or a redundancy word, and performs speech-recognition processing in a voice recognition unit according to claim 1.
[0010] Moreover, in a voice recognition unit according to claim 2, while the above-mentioned speech recognition means makes the probability value of the language model of a word connected to the pause in the above-mentioned utterance voice sentence the predetermined value which is a value near 1 or 1, when connecting with words other than a pause through a pause, it calculates the probability value of the language model of a word by skipping the above-mentioned pause, and performs speech recognition processing.
[0011] Furthermore, in a voice recognition unit according to claim 3, while the above-mentioned speech recognition means makes the probability value of the language model of a word connected to the redundancy word in the above-mentioned utterance voice sentence the predetermined value which is a value near 1 or 1, when connecting with words other than a redundancy word through a redundancy word, it calculates the probability value of the language model of a word by skipping the above-mentioned redundancy word, and performs speech recognition processing.

[Translation done.]

* NOTICES *

Japan Patent Office is not responsible for any damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

DETAILED DESCRIPTION

[Detailed Description of the Invention]
[0001]
[Industrial Application] This invention relates to the voice recognition unit which processes the pause and redundancy word in utterance voice.
[0002]
[Description of the Prior Art] In recent years, research of a continuous speech recognition is done briskly and the sentence voice recognition system is built by some research facilities. Many of these systems make applicable to an input the voice uttered carefully. however -- human beings' communication -- "that -" -- "-- obtaining - " -- etc. -- it is in the redundancy word represented and the condition (henceforth a pause) which does not have utterance voice temporarily -- it says and stagnation, a misstatement, a correction, etc. appear frequently.
[0003]
[Problem(s) to be Solved by the Invention] On the other hand, it is One to one of the algorithms which can be used for a continuous speech recognition. pass There is a DP (Viterbi search) algorithm. In the voice recognition unit using this algorithm, when a pause and a redundancy word were in the inputted utterance voice, there was a trouble that a sentence recognition rate fell.
[0004] It is in offering the voice recognition unit which can acquire a high sentence recognition rate as compared with the conventional example, even if the purpose of this invention is the case where the above trouble was solved and a pause and a redundancy word are in the inputted utterance voice.
[0005]
[Means for Solving the Problem] The voice recognition unit according to claim 1 concerning this invention The utterance voice sentence which consists of a character string inputted with reference to the predetermined statistical language model In the voice recognition unit which carries out speech recognition by calculating the probability value of the word which should be carried out speech recognition based on one piece or two or more words which are connected before the word While making the probability value of the language model of a word connected to the pause or redundancy word in the above-mentioned utterance voice sentence into the predetermined value which is a value near 1 or 1 When connecting with words other than a pause and a redundancy word through a pause or a redundancy word, it is characterized by having a speech recognition means to calculate the probability value of the language model of a word by skipping the above-mentioned pause or a redundancy word, and to perform speech recognition processing.
[0006] A voice recognition unit according to claim 2 moreover, the utterance voice sentence which consists of a character string inputted with reference to the predetermined statistical language model In the voice recognition unit which carries out speech recognition by calculating the probability value of the word which should be carried out speech recognition based on one piece or two or more words which are connected before the word While making the probability value of the language model of a word connected to the pause in the above-mentioned utterance voice sentence into the predetermined value

which is a value near 1 or 1 When connecting with words other than a pause through a pause, it is characterized by having a speech recognition means to calculate the probability value of the language model of a word by skipping the above-mentioned pause, and to perform speech recognition processing.
[0007] A voice recognition unit according to claim 3 furthermore, the utterance voice sentence which consists of a character string inputted with reference to the predetermined statistical language model In the voice recognition unit which carries out speech recognition by calculating the probability value of the word which should be carried out speech recognition based on one piece or two or more words which are connected before the word While making the probability value of the language model of a word connected to the redundancy word in the above-mentioned utterance voice sentence into the predetermined value which is a value near 1 or 1 When connecting with words other than a redundancy word through a redundancy word, it is characterized by having a speech recognition means to calculate the probability value of the language model of a word by skipping the above-mentioned redundancy word, and to perform speech recognition processing.
[0008] Moreover, a voice recognition unit according to claim 4 is characterized by the above-mentioned predetermined value being 1.0 or less [0.8 or more] in a voice recognition unit according to claim 1, 2, or 3. Furthermore, a voice recognition unit according to claim 5 is characterized by the above-mentioned statistical language model being trigram of a word in a voice recognition unit according to claim 1, 2, 3, or 4.
[0009]
[Function] While the above-mentioned speech-recognition means makes the probability value of the language model of a word connected to the pause or the redundancy word in the above-mentioned utterance voice sentence the predetermined value which is a value near 1 or 1, when connecting with words other than a pause and a redundancy word through a pause or a redundancy word, it calculates the probability value of the language model of a word by skipping the above-mentioned pause or a redundancy word, and performs speech-recognition processing in a voice recognition unit according to claim 1.
[0010] Moreover, in a voice recognition unit according to claim 2, while the above-mentioned speech recognition means makes the probability value of the language model of a word connected to the pause in the above-mentioned utterance voice sentence the predetermined value which is a value near 1 or 1, when connecting with words other than a pause through a pause, it calculates the probability value of the language model of a word by skipping the above-mentioned pause, and performs speech recognition processing.
[0011] Furthermore, in a voice recognition unit according to claim 3, while the above-mentioned speech recognition means makes the probability value of the language model of a word connected to the redundancy word in the above-mentioned utterance voice sentence the predetermined value which is a value near 1 or 1, when connecting with words other than a redundancy word through a redundancy word, it calculates the probability value of the language model of a word by skipping the above-mentioned redundancy word, and performs speech recognition processing.
[0012]
[Example] Hereafter, the voice recognition unit of the example which starts this invention with reference to a drawing is explained. In the voice recognition unit of this example of drawing 1 the phoneme collating section 4 It is based on the data about the utterance voice inputted, and is a hidden Markov model (it is hereafter called HMM.). The One DP speech recognition section after recognizing a redundancy word and a pause with reference to HMM which is a sound model in memory 5 pass (it is hereafter called the speech recognition section.) One using DP algorithm pass 6 is characterized by recognizing the voice containing a redundancy word and/or a pause by skipping a redundancy word or a pause, when performing speech recognition with reference to the statistical language model in the statistical language model memory 7 (trigram of a word).
[0013] First, amelioration of sentence voice-recognition algorithm is described. Above One pass In the path computation of DP algorithm, in order to acquire the word train of the maximum **, there are two

approaches.
(1) Trace back : in each time of day and each condition, when the maximum accumulation likelihood is calculated, memorize the selected path. And speech recognition processing is performed by performing the trace back after termination of likelihood count in accordance with the path by which storage was carried out [above-mentioned] (henceforth the 1st path computation approach). ; For example, refer to 15213 Kai-Fu Lee, "Large-Vocabulary Speaker Independent Continuous Speech Recognition:The SPHINX System", and CMU-CS-88 April 18, 1988 [-148 or]. .
(2) In :each time of day and each condition which are simultaneously calculated with the maximum accumulation likelihood, when the maximum accumulation likelihood is calculated, perform speech recognition processing by passing the path chosen as coincidence to the following condition (henceforth the 2nd path computation approach). ; For example, refer to Jin-ichi Murakami, "one algorithm of a continuous speech recognition using trigram of a word", acoustical-societies-of-america lecture collected works, pp.185-186, and 2-Q-October, 1992 [7 or]. .
[0014] Since there is little computational complexity and it ends, the path computation approach of the above 1st is often used from the former. On the other hand, although computational complexity increases as compared with the 1st path computation approach, the path computation approach of the above 2nd Since it is possible to get to know a path even if it does not act as the trace back in each time of day and each condition, It is easy to combine with LR parser of the left-right mold in a language model, and since it ends by little memory as compared with the 1st path computation approach when it is many, the speech recognition section 6 of this example adopts the 2nd latter path computation approach.
[0015] Hereafter, with reference to drawing 1 which shows the voice recognition unit using the speech recognition approach of this example, the configuration and actuation of the voice recognition unit using the statistical language model of this example are explained.
[0016] In drawing 1 , after a speaker's utterance voice is inputted into a microphone 1 and changed into a sound signal, it is inputted into the feature-extraction section 2. the LPC analysis after the feature-extraction section 2 carries out A/D conversion of the inputted sound signal -- performing -- a logarithm -- power, a 16th cepstrum multiplier, and delta -- a logarithm -- the 34-dimensional feature parameter containing power and 16th delta cepstrum multiplier is extracted. The time series of the extracted feature parameter is inputted into the phoneme collating section 4 through buffer memory 3. HMM in the hidden Markov model (henceforth HMM) memory 5 connected to the phoneme collating section 4 consists of arcs which show transition between two or more conditions and each condition, and has the transition probability between conditions, and a output probability to input code in each arc. The phoneme collating section 4 performs phoneme collating processing based on the inputted data, and outputs phoneme data to the speech recognition section 6.
[0017] The statistical language model memory 7 which memorizes beforehand the predetermined statistical language model containing trigram of a word is connected to the speech recognition section 6. The speech recognition section 6 refers to the statistical language model in the statistical language model memory 7, and is predetermined One. pass By processing without back track rightward from the left, and determining the word of a higher occurrence probability as speech recognition result data about the inputted phoneme data, using DP algorithm, processing of speech recognition is performed and the determined speech recognition result data (character-string data) are outputted.
[0018] Subsequently, processing of the pause in the speech recognition section 6 is explained. Although a pause appears between clauses in many cases, it may appear in all audio locations. However, in the conventional language model, since it cannot follow in footsteps of this, incorrect recognition tends to occur in the section of a pause. Then, the statistical language model which is a word and which becomes trigram and One pass Using DP algorithm, even if the pause was inputted into the boundary of all words and words, the speech recognition approach in which sentence recognition is possible was invented. In the example concerning this invention, a pause is first considered to be one word and the value of trigram of the word connected to a pause is set to 1.0. And when connecting with words other than a pause, trigram of a word is calculated by skipping a pause. for example, -- "-- it is called Tokyo Minato-ku

Shinbashi / pause (pause) / 1 chome" -- a character string -- when a sentence is inputted, the probability is calculated with P(Shinbashi ** Tokyo Minato-ku) x1.0xP (1 chome ** Minato-ku Shinbashi). Here, P (A|B) is the probability for the word "A" to come after the word "B", and is the same hereafter. Although it is an approximate solution when it does in this way, the solution of the maximum ** when using the word "trigram", except for a pause is acquired.

[0019] In the above example, although the value of trigram of the word connected to a pause is set to 1.0, this invention sets the value of trigram of the word connected not only to this but to a pause as the value which has or more 0.8 1.0 or less range preferably.

[0020] Drawing 2 is a flow chart which shows the trigram computation of the pause performed in the voice recognition unit of drawing 1 . In addition, it is processed like [word / redundancy] the flow of drawing 2 . The computation concerned is the approach of calculating trigram of a word w0, when the word trains w3, w2, w1, and w0 are inputted, as shown in drawing 2 , in step S1, it is judged first whether a word w0 is a pause, and it is judged in step S2 whether Tango w1 is a pause. If it is YES in step S1, it will progress to step S6, after setting the value TR of trigram of a word w0 as 1.0 in step S3. If it is [in / in NO, progress to step S2, and / step S2] YES in step S1, it will progress to step S6, after setting up a probability value P (w0|w3, w2) as a value TR of trigram of a word w0 in step S4. On the other hand, if it is NO in step S2, it will progress to step S6, after setting up a probability value P (w0|w2, w1) as a value TR of trigram of a word w0 in step S5. In step S6, the value TR of trigram of the set-up word w0 is outputted as calculated value, and the computation concerned is ended.

[0021] Furthermore, processing of the redundancy word in the speech recognition section 6 is explained. free utterance -- "that -" -- "-- obtaining - " -- etc. -- many redundancy words appear. The count of an appearance in the free utterance which this invention person collected from the test data which carries out the detail after-mentioned shows two or more redundancy words in Table 1 thru/or 3.

[0022]

[Table 1]

--------------- A redundancy word The count of an appearance --------------- "**" 604 "**-" 268 "** - **" 2 "** - **" 5 -- "-- such -- " -- 7 "****" 151 -- "-- that -- " -- 1809 "that -" 2025 -- "-- that -- obtaining -- " -- 77 -- "-- that -- obtaining -" -- 3 -- "-- it is -- " -- 26 -- "-- it is -" -- 58 "no, -" 2 -- "-- obtaining --"23 -- "-- obtaining -" -- 71 "well" -- 26 "well" 2 "it does not obtain" 7 -- "-- obtaining -- " -- 1040 -- "-- obtaining -" 3105 "** - **" 256"** - ** and -" 2 --------------- [0023]

[Table 2]

--------------- A redundancy word The count of an appearance --------------- "It is with ** - **." 8 -- "-- obtaining - " -- 466 -- "-- obtaining - and -" -- 4 "it obtains and is with -" 3 -- "-- obtaining - well -- " -- 3 "** - **" 2 "yes" 13 "****" 22 "**** -" 4 "****" 62 "**** and -" 11 "the sexagenary cycle" 47 "sexagenary-cycle -" 13 -- "-- " -- 59 "it is -" 196 "****" 2 -- "-- like this -- " -- 9 -- "-- this -- " -- 9 "this -" 4 -- ** -- **** -- " -- 4 "**" 8 "**-" 2--------------- [0024]

[Table 3]

--------------- A redundancy word The count of an appearance --------------- "****" 2 "**" 2 -- "-- the -- " -- 115 "the -" 48 -- "just for a moment" 8 "a **" 2 -- "-- it is -- " -- 61 -- "-- it is -" -- 13 -- "-- ** -- " -- 77 -- "-- ** -" -- 11 "**" 4 "****-" 2 "** - **" 2 "**" 263 "**-" 8"**** 186 -- "-- well -- " -- 176 "****" 5 "**" 2 "**" 27 ****-" 19 ****" 2--------------- [0025]" This redundancy word has the property which resembled the pause in that it may appear in all the locations of a sentence. Therefore, the same technique as processing of a pause can be used for processing of a redundancy word. That is, a redundancy word is skipped with a language model, recognizing a redundancy word in a sound model. for example, -- "-- it Tokyo Minato-ku Shinbashi obtains, and is called - and 1 chome" -- a character string -- when a sentence is inputted, the probability value is calculated with P(Shinbashi ** Tokyo Minato-ku) x1.0xP (1 chome ** Minato-ku Shinbashi). Therefore, in processing of a redundancy word, the "pause" in the flow chart shown in drawing 2 can be replaced with a "redundancy word", and can be processed similarly.

[0026] In the above example, although the value of trigram of the word connected to a redundancy word is set to 1.0, this invention sets the value of trigram of the word connected not only to this but to a redundancy word as the value which has or more 0.8 1.0 or less range preferably.

[0027] this invention person performed simulation, in order to perform evaluation by the sentence recognition rate using the voice recognition unit of this example explained above. Speech recognition simulation of a specified speaker and an unspecified speaker was carried out to language information using bigram or trigram of a word. The test data used the sentence (the so-called model conversation.) of an inquiry of the international congress which the narrator uttered. In addition, there is a pause of about 20 mses in the head and tail of a test data. Moreover, the chain probability value of trigram added and calculated the text of a test data out of these people's dialogue database in the about 12,000 texts (about 170,000 words) which are data about reservation of an international congress.

[0028] In the simulation which does not process an above-mentioned pause, while 78.6% of sentence recognition rate was acquired by the specified speaker in trigram, by the unspecified speaker, 59.5% of sentence recognition rate was acquired. On the other hand, in the simulation of the sentence recognition rate which performed processing of a pause shown in drawing 2 , in specified speaker recognition, 86.3% of higher sentence recognition rate was acquired, and the speech recognition engine performance improved. On the other hand, in unspecified speaker recognition, 60.3% of sentence recognition rate was acquired, and although the improvement in the speech recognition engine performance was not remarkable, it improved a little.

[0029] Furthermore, the sentence recognition simulation of free utterance is described. The effectiveness of processing of a redundancy word is investigated with the voice of free utterance especially here. Although the definition of free utterance changes with people, in the simulation using this example, the voice data recorded by the approach as shown below is used.

(1) reading voice data: -- the voice data which read out the text -- it is -- a redundancy word -- it hesitates to say and there is no correction.

(2) False free utterance data : it is the voice data which gave the redundancy word and read out the text. Except for a redundancy word, the above-mentioned (1) reading utterance data and the contents of utterance are the same, it hesitates to say them, and there is no correction.

(3) Free utterance data : a speaker is the voice data which memorized the text, understood the intention and spoke freely. The contents of utterance differ from the above-mentioned (1) reading utterance data. Here, it hesitates to say and there is no correction.

[0030] Speakers are not a narrator but ordinary persons. as a redundancy word -- "that -" -- "-- obtaining - " -- "-- well -- " -- etc. -- 109 kinds are defined. Only in trigram of the word of unspecified speaker recognition, speech recognition simulation carried out. Moreover, processing of an above-mentioned pause was also performed. This simulation result is shown in Table 4. While the recognition rate described 64.4% was acquired with the voice of the false free utterance which attached the redundancy word so that clearly from Table 4, 34.4% of recognition rate was acquired with free utterance voice.

[0031]

[Table 4]

Sentence recognition simulation result of free utterance --- Sentence recognition rate (%) ------------------ ----------- utterance format + pause processing + pause processing + redundancy word-processing ------------------ ----------- reading utterance 82.6% 74.8% False free utterance 26.7% 64.4% Free utterance 14.1% 34.4% ---------------------------- [0032] Therefore, processing of an above-mentioned redundancy word understands that it is effective in recognition of free utterance, when false free utterance or free utterance also takes into consideration that a recognition rate seldom falls to an effective thing and a list also in sentence recognition of reading utterance as compared with the conventional approach from the approach of only processing of a pause (82.6% -> 74.8%).

[0033] In the above example, although processing of the redundancy word in the speech recognition section 6 and processing of a pause are explained, the speech recognition section 6 may be constituted so that at least one side may be included among processing of a redundancy word, and processing of a pause.

[0034]
[Effect of the Invention] The utterance voice sentence which consists of a character string inputted with reference to the predetermined statistical language model according to this invention as explained in full detail above In the voice recognition unit which carries out speech recognition by calculating the probability value of the word which should be carried out speech recognition based on one piece or two or more words which are connected before the word While making the probability value of the language model of a word connected to the pause or redundancy word in the above-mentioned utterance voice sentence into the predetermined value which is a value near 1 or 1 Since it constituted so that the probability value of the language model of a word might be calculated by skipping the above-mentioned pause or a redundancy word and speech recognition processing might be performed when connecting with words other than a pause and a redundancy word through a pause or a redundancy word Speech recognition can be carried out in consideration of a pause or a redundancy word, by this, a recognition rate can improve sharply and the speech recognition engine performance can be raised.

[Translation done.]

* NOTICES *

Japan Patent Office is not responsible for any damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

PRIOR ART

[Description of the Prior Art] In recent years, research of a continuous speech recognition is done briskly and the sentence voice recognition system is built by some research facilities. Many of these systems make applicable to an input the voice uttered carefully. however -- human beings' communication -- "that -" -- "-- obtaining - " -- etc. -- it is in the redundancy word represented and the condition (henceforth a pause) which does not have utterance voice temporarily -- it says and stagnation, a misstatement, a correction, etc. appear frequently.

[Translation done.]

* NOTICES *

Japan Patent Office is not responsible for any damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

EXAMPLE

[Example] Hereafter, the voice recognition unit of the example which starts this invention with reference to a drawing is explained. In the voice recognition unit of this example of drawing 1 the phoneme collating section 4 the data about the utterance voice inputted -- being based -- a hidden Markov model (it is hereafter called HMM --) The One DP speech recognition section after recognizing a redundancy word and a pause with reference to HMM which is a sound model in memory 5 pass (it is hereafter called the speech recognition section.) One using DP algorithm pass 6 is characterized by recognizing the voice containing a redundancy word and/or a pause by skipping a redundancy word or a pause, when performing speech recognition with reference to the statistical language model in the statistical language model memory 7 (trigram of a word).
[0013] First, amelioration of sentence voice-recognition algorithm is described. Above One pass In the path computation of DP algorithm, in order to acquire the word train of the maximum **, there are two approaches.
(1) Trace back : in each time of day and each condition, when the maximum accumulation likelihood is calculated, memorize the selected path. And speech recognition processing is performed by performing the trace back after termination of likelihood count in accordance with the path by which storage was carried out [above-mentioned] (henceforth the 1st path computation approach). ; For example, refer to 15213 Kai-Fu Lee, "Large-Vocabulary Speaker Independent Continuous Speech Recognition:The SPHINX System", and CMU-CS-88 April 18, 1988 [-148 or]. .
(2) In :each time of day and each condition which are calculated simultaneously with the maximum accumulation likelihood, when the maximum accumulation likelihood is calculated, perform speech recognition processing by passing the selected path simultaneously to the following condition (henceforth the 2nd path computation approach). ; For example, refer to Jin-ichi Murakami, "one algorithm of a continuous speech recognition using trigram of a word", acoustical-societies-of-america lecture collected works, pp.185-186, and 2-Q-October, 1992 [7 or]. .
[0014] Since there is little computational complexity and it ends, the path computation approach of the above 1st is often used from the former. On the other hand, although computational complexity increases as compared with the 1st path computation approach, the path computation approach of the above 2nd Since it is possible to get to know a path even if it does not act as the trace back in each time of day and each condition, It is easy to combine with LR parser of the left-right mold in a language model, and since it ends by little memory as compared with the 1st path computation approach when it is many, the speech recognition section 6 of this example adopts the 2nd latter path computation approach.
[0015] Hereafter, with reference to drawing 1 which shows the voice recognition unit using the speech recognition approach of this example, the configuration and actuation of the voice recognition unit using the statistical language model of this example are explained.
[0016] In drawing 1 , after a speaker's utterance voice is inputted into a microphone 1 and changed into a sound signal, it is inputted into the feature-extraction section 2. the LPC analysis after the feature-extraction section 2 carries out A/D conversion of the inputted sound signal -- performing -- a

logarithm -- power, a 16th cepstrum multiplier, and delta -- a logarithm -- the 34-dimensional feature parameter containing power and 16th delta cepstrum multiplier is extracted. The time series of the extracted feature parameter is inputted into the phoneme collating section 4 through buffer memory 3. HMM in the hidden Markov model (henceforth HMM) memory 5 connected to the phoneme collating section 4 consists of arcs which show transition between two or more conditions and each condition, and has the transition probability between conditions, and a output probability to input code in each arc. The phoneme collating section 4 performs phoneme collating processing based on the inputted data, and outputs phoneme data to the speech recognition section 6.

[0017] The statistical language model memory 7 which memorizes beforehand the predetermined statistical language model containing trigram of a word is connected to the speech recognition section 6. The speech recognition section 6 refers to the statistical language model in the statistical language model memory 7, and is predetermined One. pass By processing without back track rightward from the left, and determining the word of a higher occurrence probability as speech recognition result data about the inputted phoneme data, using DP algorithm, processing of speech recognition is perform and the determined speech recognition result data (character-string data) are output.

[0018] Subsequently, processing of the pause in the speech recognition section 6 is explained. Although a pause appears between clauses in many cases, it may appear in all audio locations. However, in the conventional language model, since it cannot follow in footsteps of this, incorrect recognition tends to occur in the section of a pause. Then, the statistical language model which is a word and which becomes trigram and One pass Using DP algorithm, even if the pause was inputted into the boundary of all words and words, the speech recognition approach in which sentence recognition is possible was invented. In the example concerning this invention, a pause is first considered to be one word and the value of trigram of the word connected to a pause is set to 1.0. And when connecting with words other than a pause, trigram of a word is calculated by skipping a pause. for example, -- "-- it is called Tokyo Minato-ku Shinbashi / pause (pause) / 1 chome" -- a character string -- when a sentence is inputted, the probability is calculated with P(Shinbashi ** Tokyo Minato-ku) x1.0xP (1 chome ** Minato-ku Shinbashi). Here, P (A|B) is the probability for the word "A" to come after the word "B", and is the same hereafter. Although it is an approximate solution when it does in this way, the solution of the maximum ** when using the word "trigram", except for a pause is acquired.

[0019] In the above example, although the value of trigram of the word connected to a pause is set to 1.0, this invention sets the value of trigram of the word connected not only to this but to a pause as the value which has or more 0.8 1.0 or less range preferably.

[0020] Drawing 2 is a flow chart which shows the trigram computation of the pause performed in the voice recognition unit of drawing 1 . In addition, it is processed like [word / redundancy] the flow of drawing 2 . The computation concerned is the approach of calculating trigram of a word w0, when the word trains w3, w2, w1, and w0 are inputted, as shown in drawing 2 , in step S1, it is judged first whether a word w0 is a pause, and it is judged in step S2 whether Tango w1 is a pause. If it is YES in step S1, it will progress to step S6, after setting the value TR of trigram of a word w0 as 1.0 in step S3. If it is [in / in NO, progress to step S2, and / step S2] YES in step S1, it will progress to step S6, after setting up a probability value P (w0|w3, w2) as a value TR of trigram of a word w0 in step S4. On the other hand, if it is NO in step S2, it will progress to step S6, after setting up a probability value P (w0|w2, w1) as a value TR of trigram of a word w0 in step S5. In step S6, the value TR of trigram of the set-up word w0 is outputted as calculated value, and the computation concerned is ended.

[0021] Furthermore, processing of the redundancy word in the speech recognition section 6 is explained. free utterance -- "that -" -- "-- obtaining - " -- etc. -- many redundancy words appear. The count of an appearance in the free utterance which this invention person collected from the test data which carries out the detail after-mentioned shows two or more redundancy words in a table 1 thru/or a table 3.

[0022]

[A table 1]

--------------- A redundancy word The count of an appearance --------------- "**" 604 "**-" 268 "** -

" 2 " - **" 5 -- "-- such -- " -- 7 "****" 151 -- "-- that -- " -- 1809 "that -" 2025 -- "-- that -- obtaining -- " -- 77 -- "-- that -- obtaining -" -- 3 -- "-- it is -- " -- 26 -- "-- it is -" -- 58 "no, -" 2 -- "-- obtaining --"23 -- "-- obtaining -" -- 71 "well" -- 26 "well" 2 "it does not obtain" 7 -- "-- obtaining -- " -- 1040 -- "-- obtaining -" 3105 "** - **" 256"** - ** and -" 2 --------------- [0023]

[A table 2]

--------------- A redundancy word The count of an appearance --------------- "It is with ** - **." 8 -- "-- obtaining - " -- 466 -- "-- obtaining - and -" -- 4 "it obtains and is with -" 3 -- "-- obtaining - well -- " -- 3 "** - **" 2 "yes" 13 "****" 22 "**** -" 4 "****" 62 "**** and -" 11 "the sexagenary cycle" 47 "sexagenary-cycle -" 13 -- "-- " -- 59 "it is -" 196 "****" 2 -- "-- like this -- " -- 9 -- "-- this -- " -- 9 "this -" 4 -- ** -- **** -- " -- 4 "**" 8 "**-" 2--------------- [0024]

[A table 3]

--------------- A redundancy word The count of an appearance --------------- "****" 2 "**" 2 -- "-- the -- " -- 115 "the -" 48 -- "just for a moment" 8 "a **" 2 -- "-- it is -- " -- 61 -- "-- it is -" -- 13 -- "-- ** -- " -- 77 -- "-- ** -" -- 11 "**" 4 "****-" 2 "** - **" 2 "**" 263 "**-" 8"**** 186 -- "-- well -- " -- 176 "****" 5 "**" 2 "**" 27 ****-" 19 ****" 2--------------- [0025]" This redundancy word has the property which resembled the pause in that it may appear in all the locations of a sentence. Therefore, the same technique as processing of a pause can be used for processing of a redundancy word. That is, a redundancy word is skipped with a language model, recognizing a redundancy word in a sound model. for example, -- "-- it Tokyo Minato-ku Shinbashi obtains, and is called - and 1 chome" -- a character string -- when a sentence is inputted, the probability value is calculated with P(Shinbashi ** Tokyo Minato-ku) x1.0xP (1 chome ** Minato-ku Shinbashi). Therefore, in processing of a redundancy word, the "pause" in the flow chart shown in drawing 2 can be replaced with a "redundancy word", and can be processed similarly.

[0026] In the above example, although the value of trigram of the word connected to a redundancy word is set to 1.0, this invention sets the value of trigram of the word connected not only to this but to a redundancy word as the value which has or more 0.8 1.0 or less range preferably.

[0027] this invention person performed simulation, in order to perform assessment by the sentence recognition rate using the voice recognition unit of this example explained above. Speech recognition simulation of a specified speaker and an unspecified speaker was carried out to language information using bigram or trigram of a word. The test data used the sentence (the so-called model conversation.) of an inquiry of the international congress which the narrator uttered. In addition, there is a pause of about 20 mses in the head and tail of a test data. Moreover, the chain probability value of trigram added and calculated the text of a test data out of these people's dialogue database in the about 12,000 texts (about 170,000 words) which are data about reservation of an international congress.

[0028] In the simulation which does not process an above-mentioned pause, while 78.6% of sentence recognition rate was acquired by the specified speaker in trigram, by the unspecified speaker, 59.5% of sentence recognition rate was acquired. On the other hand, in the simulation of the sentence recognition rate which performed processing of a pause shown in drawing 2 , in specified speaker recognition, 86.3% of higher sentence recognition rate was acquired, and the speech recognition engine performance improved. On the other hand, in unspecified speaker recognition, 60.3% of sentence recognition rate was acquired, and although the improvement in the speech recognition engine performance was not remarkable, it improved a little.

[0029] Furthermore, the sentence recognition simulation of free utterance is described. The effectiveness of processing of a redundancy word is investigated with the voice of free utterance especially here. Although the definition of free utterance changes with people, in the simulation using this example, the voice data recorded by the approach as shown below is used.

(1) reading voice data: -- the voice data which read out the text -- it is -- a redundancy word -- it hesitates to say and there is no correction.

(2) False free utterance data : it is the voice data which gave the redundancy word and read out the text. Except for a redundancy word, the above-mentioned (1) reading utterance data and the content of

utterance are the same, it hesitates to say them, and there is no correction.
(3) Free utterance data : a speaker is the voice data which memorized the text, understood the intention and spoke freely. The content of utterance differs from the above-mentioned (1) reading utterance data. Here, it hesitates to say and there is no correction.
[0030] Speakers are not a narrator but ordinary persons. as a redundancy word -- "that -" -- "-- obtaining - " -- "-- well -- " -- etc. -- 109 kinds are defined. Only in trigram of the word of unspecified speaker recognition, speech recognition simulation carried out. Moreover, processing of an above-mentioned pause was also performed. This simulation result is shown in a table 4. While the recognition rate described 64.4% was acquired with the voice of the false free utterance which attached the redundancy word so that clearly from a table 4, 34.4% of recognition rate was acquired with free utterance voice.
[0031]
[A table 4]
Sentence recognition simulation result of free utterance --- Sentence recognition rate (%)
------------------ ----------- utterance format + pause processing + pause processing + redundancy word-processing ------------------ ----------- reading utterance 82.6% 74.8% False free utterance 26.7% 64.4% Free utterance 14.1% 34.4% ------------------------------ [0032] Therefore, processing of an above-mentioned redundancy word understands that it is effective in recognition of free utterance, when false free utterance or free utterance also takes into consideration that a recognition rate seldom falls to an effective thing and a list also in sentence recognition of reading utterance as compared with the conventional approach from the approach of only processing of a pause (82.6% -> 74.8%).
[0033] In the above example, although processing of the redundancy word in the speech recognition section 6 and processing of a pause are explained, the speech recognition section 6 may be constituted so that at least one side may be included among processing of a redundancy word, and processing of a pause.

[Translation done.]